

Research Search Tool

Chris Hesser
chesser3@gatech.edu

Abstract— A novel website-based tool for exploring research is presented. Inspired by random restart hill climbing, the tool helps users find peaks of interest in research areas by presenting a pseudo-random article from the CORE open access research collection then allowing users to quickly narrow their search space by selecting keywords from works that interest them. The goals of this tool are primarily introspective, helping users discover research areas about personally interesting topics in a timely way and aiming to solve the problems associated with not selecting areas of research due to a lack of inspiration.

1 INTRODUCTION

Students pursuing research with the intent of providing a novel insight, tool, or idea are first challenged by selecting an area to research. This process is complex and one must weigh many factors such as scope, advisor input, personal interest, time, and feasibility when choosing an area to explore (Barr, 1984). Additionally, the amount of published and submitted research is growing (Jinha, 2010) making it difficult to go where no one has gone before.

Tools such as Google Scholar are invaluable during the research phase, but fail to provide the seed of inspiration. Somewhere to start and something to judge are inherently not provided by the homepage of Google scholar.

This paper suggests a novel approach to finding, exploring, and selecting interesting areas of research, a bottom up approach wherein researchers are given a pseudo-random article to “judge”. Judgement occurs by selecting keywords from the article title and requesting a new resource from a narrowed space defined by the chosen keywords. Another resource is presented and judged, and another, until articles appear that align with the user’s interest... or they discover they don’t like where they are going. Restarting the process allows for exploring a completely different area.

Eventually after quickly navigating different research areas, the goal is that the user has formed an opinion on what area of research is appropriate for their needs and have done so in a timely manner. This is strictly a goal since one cannot possibly hope for enough time to explore all areas of research, but as with any difficult optimization problem, a decent solution found on time is better than taking an eternity to search for the best one.

2 RELATED WORK

Tools for aiding research are not new, but the existing tools are not designed for discovery and exploration of unknown (from the researcher's perspective) topics. One such tool is Questia (Questia, 2020). Geared at undergraduate learners, Questia offers a library of liberal arts books and articles. One neat feature is a slot machine styled topic generator which randomly gives a keyword which if selected offers resources related to that keyword. While the tool is great for writing a history paper, its library does not offer enough resources for those looking to contribute something new.

Another way to search for areas of research is to use a search engine. Already mentioned is the powerhouse Google Scholar, but most online journals offer some way to query for articles.. One of the powerful features of search engines is the ability to show results in order of relevance. This makes it easy to find sources that others have found useful and make a researcher aware of the most impactful works quickly. However, one critical input is needed: a query. This must come from the researcher. But how can a query describing something unknown come from the unknowing researcher? The answer is they need to be inspired by an outside source.

One can absolutely use a search engine to become inspired, starting at a base topic and then refining the search. But this is time consuming and while the first page may show the most relevant articles for a keyword, they are likely related to one another giving little chance that those resources provide the researcher a broad understanding of all existing research areas.

Since the primary shortcomings of search engines stem from users not knowing what they don't know and not describing a research space, a tool that overcomes those problems prove helpful to researchers.

To start, how can a tool provide inspiration to researchers? What is inspiration anyway? Thrash and Elliot (2004) set out to answer the latter question in one of many articles on inspiration. They identify that inspiration has three components: evocation, transcendence, and motivation. Evocation here means that inspiring thoughts enter unwillingly or without one's control. Transcendence means that one has grown to consider views that were not concerning before. The last characteristic, motivation moves one to do something. However, the word inspired is used in two different ways, inspired by something or inspired to do something (Hesser, 2020).

Additionally, Thrash et al (2014) differentiate between typical usages of inspiration as a word. Primarily inspiration can refer to "*inspired by*" or "*inspired to*". "Inspired by" refers to being inspired by something. This usage implies that the first two parts of inspiration are present. Using "inspired to" in language implies that all three components are present in the user. Since a researcher needs to actualize their inspiration in thought deed or long paper, we can use "inspired to" to describe the goals of a research tool.

As far as the intended user experience, inspiration requires the user to be in an accepting state, the tool itself would need to place as many users as it can in such a state. A user's affective state on interacting with a website can be affected by website design. Color is one such example. Wilms and Oberfield (2018) performed a study where subjects are placed in a room with a tv with a certain color, asked to do an activity with the tv, and their physical response was measured via sensors on the body and a questionnaire. They found that arousal is associated with higher brightness and saturation. Other studies further imply that color of text, color of background, location of text on mobile devices, and color schemes also have an impact on an individual's emotions (Heidig et al., 2015; Hall & Hanna, 2004; Jiang et al., 2019). In another study, decorative pictures placed on websites were found to increase positive emotions in website users (Schneider et al., 2016). Based on this research, color choices, placement of text, and overall aesthetics do have a large impact on the emotional state of website users. By invoking positive affect, it is possible to design a website that fosters the emotions needed for inspiration (Hesser, 2020).

The tool has to look good and fulfill users' needs in order to realize the intended user experiences. In order to fulfill a user's needs, the user must be identified and in this case the chosen end users are researchers intending to contribute

something new, primarily graduate level and above researchers. With these users in mind, the tool should facilitate fast exploration, help describe research spaces, and inspire users.

The suggested way to provide for all of these needs is to follow the analogy of Random Restart Hill Climbing (Forrest and Mitchell, 1993). This particular algorithm is normally applied to optimization problems where attempting to find the optimum solution through brute force or guaranteed optimum producing methods is too time consuming (or expensive, whatever that means). Random Restart Hill Climbing (RHC) works by randomly jumping to some point in the space evaluating the value of that point and the points around it, and then moving to the next point to repeat this process until the no points around it are better. This is the top of a hill metaphorically, a local optimum, and possibly the global optimum over the space. This process is repeated and the peaks are compared with each other to find the best one with the hope that this randomized process finds the best or at least a good enough solution.

Some key points to RHC being effective include having a fast point evaluation function and having a “nicely” sloping space. Fast point evaluation allows for the algorithm to quickly climb peaks, allowing for more efficient exploring. Nicely sloping spaces implies that the space does not have very sharp narrow peaks surrounded by sub optimum peaks because landing in a spot conducive to climbing the optimum hill would require the peak itself or a very small space around it be randomly chosen

In order to implement this analogy as a research tool, the space of research needs to be described in a granular way, there needs to be a way to evaluate whether one point in the space is better than others, the space needs to behave in a way that is conducive to RHC, and the evaluations need to happen quickly so as to be efficient with a user’s limited time.

3 SOLUTION

In order to realize the goals described above the following is proposed: an emotionally affective website providing random article titles to evaluate wherein the evaluation limits the space from which the next random article is chosen. Evaluation is done by selecting words of interest from a title that must be included by subsequent resource content.

Drawing the analogies to RHC, points in the space are represented by research articles, point evaluation is done by the user via selecting words from the resource, and peaks occur when the user cannot find a better space by selecting more words and must restart in order to explore more. Additionally, the key functional needs of RHC are preserved. Research articles represent a small yet arguably comprehensive representation of the space of research which can be evaluated for value quickly. The space as described by research articles behaves in a way that is conducive to RHC since user interest in a paper is likely to be similar to nearby papers in the space (ie. gently sloped).

Since the target audience is possibly lacking inspiration, the evaluation of one research article compared to another might have no slope, a big plateau of no interest. Because of this the tool looks to instill a positive affect in the user by using UI design techniques to make the tool fun, useful, efficient, and aesthetically pleasing. Doing this will help users to make their evaluations more relevant, causing peaks of interest to be formed out of the plateau. Specifically, this is done by using an image, animations, and a streamlined flow focussed on a balance of limiting information overload and allowing for quick searching

3.1 UI implementation

For streamlining the search process the tool needs to be intuitive, but the user also needs to learn how to use the tool with high efficiency to take full advantage and this would take a long time to understand through trial and error learning. To solve this problem a short how-to video is shown on the landing page describing how to use the interface in a fast way. After navigating away from the landing page, users are shown the main search page.

The main search page (Figure 1) is a simple interface offering a search bar, a reset button, a next button, a home button, a read description button, and a resource section. The resource section is the main feature and presents the research resource title, author names, and date. When hovering over a word in the title, an individual word changes color which signals that it is an element the user can interact with. Upon clicking the word it is added to the search bar. Users can also type into the search bar. The next button fetches a new article based on the words in the search bar. The reset button signifies the top of a peak in the RHC analogy which clears the search bar and fetches a new article. The home button navigates

back to the home page and the read description button navigates users to the description page.

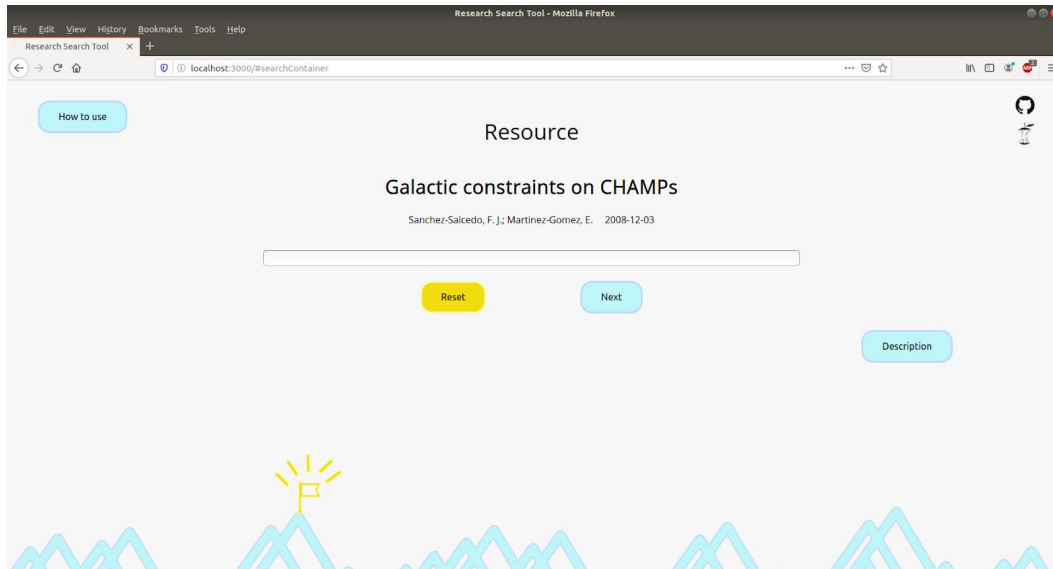


Figure 1 - Search page showing a resource title (Sanchez-Salcedo Martinez-Gomez, 2008) using CORE's API.

The description page shows a short description of the resource if available, a link to the full text or pdf, and a button to navigate back to the search screen. Users that have reached this screen are reaching more interesting areas of their search where it is feasible to spend time reading something other than the article title. Users who find themselves navigating to the full resources frequently are possibly nearing the end of the tools intended usage. This signals that they should move to a more focused querying tool such as Google Scholar.

3.2 Article fetching

Presenting pseudo random articles is the primary feature of this tool and this is implemented through a middleware type backend service. The backend service exercises an API from an organization called CORE (this is not to be confused with core api which is a web development tool) (CORE, 2020). CORE offers access, data mining, and querying of a large collection of open access research articles. The middleware backend service simply prepares queries from the front end, sends it to CORE, and returns a single resource back to the front end to display.

3.3 Pseudo random resource generation

One design goal of the initial article is that it is a random resource. CORE's API does not offer truly random access to resources without downloading the entire collection (a future enhancement possibly), but it does allow for querying with '*' which means search for everything. It's unclear how resources are ordered in a search like this, but in theory any resource can be returned by this query. Additionally, one can request which page number from the search to receive (just like a page in a Google search). To wrap everything together, the backend queries for the i'th page number using a query of '*' where i is a random number from 1 to 100. Each page contains 10 resources and the j'th one is sent back to the front end where j is a random number from 1 to 10.

When a user presents a query other than empty, the query needs to reduce the space from which articles are sampled. The default interface from CORE expands the space as words are added, so a '+' is added before each word so that the space is limited with each additional word. This does imply that users can receive zero resources back from their query depending on their chosen combination of words. However when users click words from the title, empirical evidence shows users tend to have sufficient related resources to access likely because the language in titles is used in a number of article titles or descriptions.

4 METHODOLOGY

The scope of this article is mostly to present a design, but it is worth discussing metrics that might be valuable in assessing the tools effectiveness in the future. User satisfaction surveys would be one way to do this. Since users are the only ones who can understand if they are becoming inspired or getting a more focussed research direction, surveys would be a helpful way to assess those. Another potential way to assess usefulness of the tool is to analyze clicking metrics from the page and time spent using the page. Ideally, users will find their inspiration quickly and then leave the page. From a user flow perspective, the backend could maintain button click metrics and also how long users are spending on the page. Comparing search and reset clicks with description page navigation or full resource clicks could for example help show if resource titles are helpful enough for users to assess their usefulness. High navigation to the description page would signal that users are either not getting enough

information from the title or that they should leave the page and do a more focussed search.

As far as deciding which image should be used, a survey was done of graduate students asking them to choose between two different themes, one a mountain theme related directly to the analogy of RHC and the other a more artistic flower graphic (Figure 2). Then they were asked to rate readability, utility and disloyalty towards their chosen theme (Figure 3). Overall, The mountain theme was preferred by students and had lower disloyalty associated with the image. Curiously, the flower image had higher ratings of clickability and readability possibly showing that the users were happier to defend their choice.

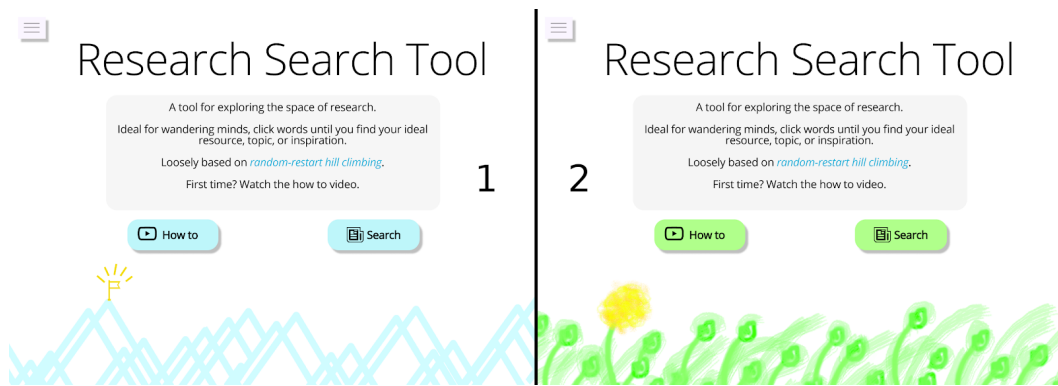


Figure 2 - Image of survey choices presented to students. Mountain theme is one the left (1), flower theme on the right (2)

Theme	Count	Ave readability	Ave clickability	Ave disloyalty
mountains	41	3.50	3.48	2.62
flowers	24	3.76	4.08	2.76

Figure 3— Survey results of mountain vs flower theme. Responses are on a 1 to 5 scale, 5 being the highest value..

5 RESULTS

The website is not currently hosted, but the source code of the project is available here: <https://github.com/chris-hesser/ResearchTool>

Initial usage of the tool looks promising, providing a fun activity for searching, an aesthetically pleasing design (for mountain folk at least), and a focussed set of pages for displaying information.

6 LIMITATIONS

The largest misconception of the tool is the user's idea of searching and what that entails. We are used to focused search tools such as Google where only relevant or top results are seen and only from queries that users provide. The tool proposed by this paper does not allow for an easy way to pinpoint articles due to the random nature of the tool and the design purposefully not returning the most relevant results. This misconception is mitigated primarily through the landing page how-to video which lets users know in under a minute what to expect from the tool.

Another limitation is the choice of research articles to describe research. While serving the needs of this tool in general, research articles in themselves do not describe all areas of thought exploration or academic research. Underrepresented areas of research are similarly underrepresented in the results of the tool. In the same way that the sample space is skewed unevenly depending on the amount of research existing in an area, resources themselves can contain varying amounts of information and a title itself does not represent the amount of knowledge contained in a resource. This implies that there can be many potential research topics or points represented by a resource that cannot be exploited by this tool as it is now.

One last note is that most of the resources from the CORE website are in English. This limits both the usage of the tool to english speakers and also limits the range of research covered by the tool.

7 CONCLUSION

The eventual goal of this research tool is to be hosted online for researchers to do exploring and find inspiration, but for now this project is a useful framework to

realize the full vision. As a representation of RHC, the tool provides the infrastructure to search the space of research, provide inspiration, and does so in a fast, easy to use manner. The page aesthetics allow the user to become inspired and the usage of CORE's API allows the tool to portray a general representation of existing research. The presented pseudo random algorithm for exercising CORE's API allows users to discover resources and interests that are not previously known, which is another basis of inspiration.

8 FUTURE WORK

An obvious next step would be to host the tool and allow access to researchers around the globe. I hope to do this and will add the url to the code repository if and when that occurs.

While the tool itself is focussed on exploring research, the application of using human interest as a scoring function in a RHC style exploration may have other applications. One simple example is deciding what to eat for dinner. Instead of looking at labels such as Mexican food or Asian food, one could start with a "random" meal. Next, change the most undesirable ingredient to something better, anything better. Continue until all the ingredients are as good as they can be and then restart with a different random meal. Eventually, you will have visited many potential options for dinner and can pick the favorite.

Another potential use for the framework is as a research tool especially related to interest in topics. The UI is very simple and allows for easy customization. For example, one could use the tool to gauge how a particular website graphic impacts inspiration, happiness, learning, openness to new ideas, or time using the tool. Similar tests can be done by changing color themes, button language, or title language. Modifying the backend api could allow for potential uses as a marketing research tool.

Another interesting test in measuring the tool's effectiveness could be for measuring diversity of ideas in a classroom where the professor asks students to write a research paper.

While the RHC serves as an easy to implement analogy for this tool, other optimization algorithms can be explored as well. An applied genetic algorithm for example could help researchers discover useful collaborations in research and discover points where research in different fields intersects.

9 REFERENCES

1. Barr, M. A. (1984). The selection of a dissertation topic: Elements influencing student choice.
2. CORE – Aggregating the world’s open access research papers. (2020). CORE. <https://core.ac.uk/>
3. Forrest, S., & Mitchell, M. (1993). Relative building-block fitness and the building-block hypothesis. In *Foundations of genetic algorithms* (Vol. 2, pp. 109-126). Elsevier.
4. Hall, R. H., & Hanna, P. (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & information technology*, 23(3), 183-195.
5. Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior*, 44, 81-95.
6. Jiang, Z., Yin, H., Luo, Y., Gong, J., Yang, Y., & Lin, M. (2019, October). Quantitative Analysis of Mobile Application User Interface Design. In *2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC)* (pp. 1-8). IEEE.
7. Jinha, A. E. (2010). Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, 23(3), 258-263.
8. Hesser, C (2020). Qualifier Question.
9. Sanchez-Salcedo, F. J., & Martinez-Gomez, E. (2008). Galactic constraints on CHAMPs. *arXiv preprint arXiv:0812.0797*.
10. Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction*, 44, 65-73.
11. Thrash, T. M., & Elliot, A. J. (2004). Inspiration: core characteristics, component processes, antecedents, and function. *Journal of personality and social psychology*, 87(6), 957.

12. Thrash, T. M., Moldovan, E. G., Oleynick, V. C., & Maruskin, L. A. (2014). The psychology of inspiration. *Social and Personality Psychology Compass*, 8(9), 495-510.
13. Questia, Your Online Research Library. (2020). Questia. <https://www.questia.com/writing-center>
14. Wilms, L., & Oberfeld, D. (2018). Color and emotion: effects of hue, saturation, and brightness. *Psychological research*, 82(5), 896-914.